

Application of computational biophysics and bioinformatics to multiscale biological problems

(Zastosowanie biofizyki obliczeniowej i bioinformatyki w wielkoskalowych problemach biologicznych)

Promotorzy:

prof. dr hab. Paweł Golik (*Instytut Genetyki i Biotechnologii*, Wydział Biologii, Uniwersytet Warszawski)

prof. dr hab. Marek Niezgódka (*Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego*, Uniwersytet Warszawski)

Recenzenci:

dr hab. Joanna Trylska (*Centrum Nowych Technologii*, Uniwersytet Warszawski)

dr hab. Krzysztof Pawłowski (*Katedra Doświadczalnictwa i Bioinformatyki*, Wydział Rolnictwa i Biologii, Szkoła Główna Gospodarstwa Wiejskiego)

Praca wykonana w ramach Międzywydziałowych Interdyscyplinarnych Studiów Doktoranckich w zakresie nauk Matematyczno-Przyrodniczych (MISDoMP).

Ostatnie dekady rozwoju nauki pokazują, jak kluczowe jest zastosowanie metod teoretycznych i obliczeniowych w naukach biologicznych, pomimo związanych z nimi przybliżeń i uproszczeń. Szczególnie jest to wyraźne w przypadku sekwencjonowania genomowego, które dostarcza nam ogromnych ilości danych, doprowadzając do sytuacji, w której ich analiza, ze względu na swoją złożoność, staje się „wąskim gardłem” wielu projektów naukowych. Innym przykładem jest biologia strukturalna, w której coraz większe zastosowanie znajdują metody dynamiki molekularnej, pozwalające badać procesy wciąż niedostępne dla metod eksperymentalnych.

W ramach mojej pracy doktorskiej wykorzystałem metody bioinformatyki oraz biofizyki obliczeniowej do badania zagadnień dotyczących podstawowych aspektów funkcjonowania komórki. Pierwszym było określenie oraz pełne scharakteryzowanie, zarówno pod względem strukturalnym jak i funkcjonalnym, zestawu metylotransferaz kodowanych w genomie drożdża *Saccharomyces cerevisiae*. Drugim była weryfikacja dotychczasowego modelu oddziaływania białko-białko, w kontekście wiązania niekowalencyjnego ubikwityny do ligandów białkowych, a w konsekwencji zaproponowanie nowego modelu opisującego ten proces.

Metylotransferazy (MTazy) stanowią bardzo liczną, lecz wciąż nie w pełni scharakteryzowaną klasę enzymów występujących we wszystkich organizmach. Są one odpowiedzialne za przeprowadzanie reakcji przeniesienia grupy metylowej z jej donora, którym jest najczęściej S-adenozyl-L-metionina (znana jako AdoMet lub SAM) na nukleofilowy akceptor, którym jest przeważnie atom azotu bądź tlenu w cząsteczce białka lub kwasu nukleinowego. MTazy biorą udział w wielu kluczowych procesach, takich jak: regulacja ekspresji genów, naprawa DNA czy biosynteza metabolitów. Ponadto metylacja jest drugą po fosforylacji najczęściej występującą modyfikacją posttranslacyjną białek, pełniącą funkcję regulatorową i sygnałową. Pomimo ogromnego znaczenia tej reakcji dla funkcjonowania komórki wciąż nie znamy pełnego zestawu metylotransferaz kodowanych w genomie jednego, dowolnego organizmu. W swojej pracy dokonałem dogłębnej charakterystyki tej klasy enzymów u eukariotycznego organizmu modelowego jakim są drożdże *Saccharomyces cerevisiae*.

Wykorzystując techniki modelowania homologicznego, ze szczególnym uwzględnieniem wyszukiwania odległych homologów za pomocą metaprofilu sekwencyjnych, wykazałem, iż metylotransferom (zbiór MTaz kodowanych w genomie danego organizmu) u *S. cerevisiae* składa się z 86 MTaz. Pośród nich 53 zostały już zweryfikowane eksperymentalnie, potwierdzono ich funkcję jak i określono specyficzność substratową. Kolejne 32 znamy tylko na podstawie poprzednich badań bioinformatycznych, w których wykazano ich homologię do znanych metylotransferaz, wymagają one jednakże wciąż potwierdzenia eksperymentalnego i określenia specyficzności. Ponadto opisałem jedną całkowicie nową metylotransferazę (Yil096cp). Wszystkie te białka, w oparciu o zbudowane przeze mnie modele struktury przestrzennej, sklasyfikowałem strukturalnie na 9 klas, odpowiadających innym zwojom, z najliczniej reprezentowanym zwojem Rossmanna.

Dodatkowo przeanalizowałem architekturę domenową dla wszystkich MTaz, znajdując kilka do tej pory nie opisanych domen, w kontekście badanych białek, które mogą przyczynić się do lepszego zrozumienia funkcji tych metylotransferaz i ich powiązania z innymi procesami w komórce.

W oparciu o dane dotyczące ekspresji genów w cyklu metabolicznym (Yeast Metabolic Cycle, YMC) wskazałem, że większość genów MTaz podlega cyklicznej ekspresji w YMC. Ponadto zaobserwowałem, że MTazy o identycznej specyficzności charakteryzują się podobnym wzorem ekspresji. W oparciu o podobieństwo wzoru ekspresji, punkt izoelektryczny, rodzaj zwoju oraz lokalizację białka w komórce opracowałem nową metodę przewidywania specyficzności substratowej MTaz. Metoda ta wykorzystuje drzewa decyzyjne wygenerowane w oparciu o powyższe własności 53 metylotransferaz o znanej specyficzności substratowej. Stosując opracowaną przeze mnie metodę przewidziałem specyficzność dla 24 z 33 niescharakteryzowanych MTaz, a moje przewidywania zostały potwierdzone dla 8 z 12 zweryfikowanych eksperymentalnie do tej pory białek, w tym i dla Yil096cp. Ponadto zaproponowany przeze mnie algorytm, wykorzystujący drzewa decyzyjne i dostępne dane o ekspresji, lokalizacji jak i fizyko-chemii białek, można zastosować w badaniu innych klas enzymów.

Wynikiem tej części pracy była charakterystyka strukturalna i funkcjonalna zbioru metylotransferaz kodowanych w genomie *S. cerevisiae*. Uzupełniona została nowatorską metodą przewidywania specyficzności substratowej, przy pomocy której oszacowałem, iż metylacja u tego organizmu jest przeprowadzana przez 35 metylotransferaz RNA, 31 metylotransferaz białkowych, 9 MTaz małych cząsteczek oraz 3 MTazy lipidów; dla pozostałych 8 białek wciąż nie znamy substratu.

W drugiej części mojej pracy doktorskiej zająłem się badaniem oddziaływania białko-białko w kontekście niekowalencyjnego wiązania ubikwityny.

Istnieją dwa główne modele opisujące oddziaływanie białko-białko.

Pierwszym z nich jest wymuszone dopasowanie („induced fit”, IF). W obrębie tego modelu, oddziaływanie pomiędzy ligandem a białkiem wymusza ich wzajemne dopasowanie konformacyjne. W takim przypadku białko w kompleksie przyjmuje konformację niedostępną dla niego w stanie niezwiązanym.

Drugim modelem jest selekcja konformacyjna („conformational selection”, CS). Model ten kładzie nacisk na dynamikę białka, wskazując, iż w stanie niezwiązanym białko nie przyjmuje stale tylko jednej konformacji, ale fluktuuje pomiędzy różnymi stanami z różnymi prawdopodobieństwami. W trakcie tych fluktuacji czasami może znaleźć się w stanie, którego konformacja jest bardzo podobna do tej, którą to białko przyjmuje w kompleksie z innym. Pod wpływem oddziaływania z ligandem zmianie ulega rozkład prawdopodobieństwa z jakim dostępne są różne stany dla białka, gdyż stabilizuje ono konformację związanego białka, która to staje się tą dominującą. W tym modelu, w przeciwieństwie do modelu IF, białko w kompleksie przyjmuje konformację, która jest dostępna dla niego także w stanie niezwiązanym.

Model CS był ostatnio wykorzystany do wyjaśnienia oddziaływania białko-białko w przypadku ubikwityny i jej białkowych ligandów. Ubikwityna pełni kluczową rolę w wielu procesach biologicznych, w szczególności jako cząsteczka sygnałowa, która np. jest markerem kierującym białka do degradacji w proteasomie.

Celem mojej pracy była weryfikacja tego modelu w oparciu o dostępne struktury kompleksów ubikwityny (zarówno krystaliczne jak i NMR-owskie), za pomocą analizy *in silico* lokalnych zmian w strukturze ubikwityny. Przeprowadziłem tę analizę poprzez porównywanie struktury związanej i niezwiązanej ubikwityny, sprawdzając zakres stosowalności modelu wymuszonego dopasowania i selekcji konformacyjnej w procesie rozpoznawania ubikwityny. Gdyby struktura związanej i niezwiązanej ubikwityny była taka sama oznaczałoby to, że wiązanie odbywa się poprzez model CS, w przeciwnym wypadku do opisu należałoby wykorzystać model IF. Porównując fragmenty struktur (ubikwityny związanej i niezwiązanej) w funkcji odległości od miejsca wiązania ligandu wykazałem, iż wokół tego miejsca w ubikwitynie, pod wpływem wiązania, dochodzi do lokalnych zmian strukturalnych, które są bardziej znaczące niż globalne zmiany struktury ubikwityny. Analiza statystyczna wykazała również, że rozkład tych lokalnych zmian jest zupełnie inny i nie może być wyjaśniony tylko poprzez model dopasowania konformacyjnego. Oznacza to, że CS nie oddaje całego mechanizmu

oddziaływania białko-białko w przypadku ubikwityny i należy go jeszcze uzupełnić o model IF, w szczególności wokół miejsca wiązania ligandu.

Wynikiem ostatniej części niniejszej rozprawy było zaproponowanie nowego modelu oddziaływania białko-białko dla ubikwityny, który uzupełnia dotychczasowy model selekcji konformacyjnej o model następującego po niej lokalnego wymuszonego dopasowania. Ponadto opracowana przeze mnie metodyka może być wykorzystana w podobnych badaniach nad innymi białkami, co umożliwi pełniejsze zrozumienie ich funkcji.